

Calcul différentiel, courbes et surfaces

Licence de Mathématiques
Université Lyon 1

Filippo Santambrogio

ce poly se base partiellement sur le poly de Calcul Différentiel et Analys Complexe de Dragoş Iftimie

Table des matières

1	Calcul différentiel	2
1.1	Définitions et premières propriétés	2
1.2	Le cas de la dimension finie	4
1.2.1	Dérivées partielles	4
1.2.2	Matrice jacobienne et gradient	5
1.2.3	Continuité des dérivées partielles et différentiabilité	5
1.3	Inégalité des accroissements finis	6
1.4	Dérivées d'ordre deux et supérieur	7
1.5	Théorème d'inversion locale	9
1.6	Théorème des fonctions implicites	10
2	Courbes et surfaces paramétrées	11
2.1	Définitions et représentation implicite	11
2.2	Courbes paramétrées : longueur, courbure et torsion	12
2.2.1	Longueur et paramétrisation par longueur d'arc	12
2.2.2	Allure d'une courbe plane	13
2.2.3	Remarques sur les courbes gauches	14
2.3	Géodésiques sur la sphère et sur d'autres surfaces	14
2.4	Première et seconde formes fondamentales	15
3	Éléments d'optimisation	17
3.1	Existence du minimum	17
3.2	Conditions nécessaires d'optimalité	17
3.3	Fonctions convexes	19
3.4	Algorithmes itératifs	20

1 Calcul différentiel

Soient $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux espaces vectoriels normés sur \mathbb{R} ou \mathbb{C} . Pour simplifier la notation, on ne mettra pas d'indice à la norme $\|\cdot\|$ lorsqu'il n'y a pas de risque de confusion. Dans le cas de \mathbb{R}^n on utilisera toujours la norme euclidienne $\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

1.1 Définitions et premières propriétés

Définition 1.1. Soit U un ouvert de E , $a \in U$ et $f : U \rightarrow F$. On dit que f est différentiable en a s'il existe une application linéaire et continue $L : E \rightarrow F$ telle que

$$f(a+h) = f(a) + L(h) + o(\|h\|) \quad \text{quand } h \rightarrow 0.$$

(On a utilisé la notation habituelle $X = o(\|h\|)$ ssi $\frac{X}{\|h\|} \rightarrow 0$ quand $h \rightarrow 0$.)

Remarques.

- Pour pouvoir étudier la différentiabilité d'une fonction en un point il faut que la fonction soit définie au voisinage de ce point.
- La notion de différentiabilité ne change pas quand on remplace les normes de E et F par des normes équivalentes.
- En dimension finie le théorème de Riesz affirme que toutes les normes sont équivalentes. Par conséquent, si E et F sont de dimension finie alors la notion de différentiabilité ne change pas quand on change les normes de E et F .

Proposition 1.2. L'application linéaire et continue L qui apparaît dans la définition 1.1 est unique. On appelle L la différentielle de f au point a et on note $L = Df(a)$ ou encore $L = f'(a)$ lorsqu'il n'y a pas de risque de confusion avec la dérivée usuelle.

Proposition 1.3. Une fonction différentiable en a est continue en a .

Proposition 1.4. Si $E = F = \mathbb{R}$, alors une fonction f est différentiable en un point a si et seulement si elle est dérivable. De plus, la différentielle $Df(a)$ est l'application linéaire et continue donnée par la multiplication par la dérivée $f'(a)$:

$$Df(a)(h) = hf'(a).$$

Exemples.

- Toute application linéaire et continue entre deux espaces normés est différentiable en tout point et sa différentielle en un point arbitraire est elle-même.
- L'application $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = x_1x_2$ est différentiable en tout point et sa différentielle est donnée par

$$Df(a)(h) = a_1h_2 + a_2h_1.$$

- Soient E_1, E_2 et F des espaces normés, $B : E_1 \times E_2 \rightarrow F$ une application bilinéaire et continue (i.e. $\|B(x_1, x_2)\| \leq C\|x_1\|\|x_2\|$). Alors B est différentiable en tout point et $DB(a)(h) = B(a_1, h_2) + B(h_1, a_2)$.

- Soit E l'espace des fonctions continues sur $[0, 1]$ à valeurs dans \mathbb{R} muni de la norme $\|\cdot\|_\infty$.

L'application $T : E \rightarrow \mathbb{R}$, $T(f) = \int_0^1 f^2$ est partout différentiable et $DT(f)(h) = 2 \int_0^1 fh$.

Nous avons que la composition de deux applications différentiables est différentiable et la différentielle de la composition est la composition des différentielles.

Proposition 1.5. Soient E, F et G trois espaces normés, U un ouvert de E , V un ouvert de F , $f : U \rightarrow F$ et $g : V \rightarrow G$. On suppose que f est différentiable en $a \in U$, que $f(a) \in V$ et que g est différentiable en $f(a)$. Alors $g \circ f$ est différentiable en a et

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a).$$

Remarque. Cet énoncé concernant la composition peut être utilisé avec l'exemple b) ci-dessus pour prouver que le produit de deux fonctions différentiables (à valeur dans \mathbb{R}) est aussi différentiable et retrouver la formule usuelle pour sa différentielle. D'autre part, la somme de deux fonctions différentiables (à valeur dans un même espace vectoriel normé arbitraire) est aussi différentiable et sa différentielle est la somme des deux différentielles, ce qui prouve qu'on peut finalement traiter toutes les fonctions construites à partir des fonctions usuelles.

Définition 1.6. Soit U un ouvert de E , $a \in U$, $v \in E$ et $f : U \rightarrow F$. On dit que f admet au point a une dérivée directionnelle dans la direction v , et on la note par $\frac{\partial f}{\partial v}(a)$, si la limite suivante existe :

$$\frac{\partial f}{\partial v}(a) = \lim_{\varepsilon \searrow 0} \frac{f(a + \varepsilon v) - f(a)}{\varepsilon}.$$

En d'autres termes, la dérivée directionnelle $\frac{\partial f}{\partial v}(a)$ est la dérivée à droite en 0 de la fonction $t \mapsto f(a + tv)$.

Proposition 1.7. Si f est différentiable en a alors f admet des dérivées directionnelles en a suivant toute direction et nous avons de plus que

$$\frac{\partial f}{\partial v}(a) = Df(a)(v).$$

Exemples.

- a) L'existence des dérivées directionnelles suivant toute direction n'entraîne pas forcément la différentiabilité de la fonction. Ni même la continuité. La fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$f(x, y) = \begin{cases} 1 & \text{si } y \geq x^2 \\ 1 & \text{si } y \leq 0 \\ 0 & \text{sinon} \end{cases}$$

admet des dérivées directionnelles en 0 qui sont nulles en toute direction. Mais f n'est pas continue en 0, et *a fortiori* n'est pas différentiable en 0.

- b) Même si la fonction est continue et que toutes ses dérivées directionnelles existent en un point cela n'implique toujours pas que la fonction est différentiable. La fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$f(x, y) = \begin{cases} \frac{x(x^2 - 3y^2)}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

est continue en 0 et toutes ses dérivées directionnelles en 0 existent :

$$\frac{\partial f}{\partial v}(0) = f(v) \quad \forall v \in \mathbb{R}^2.$$

Elle n'est cependant pas différentiable en 0 car les dérivées directionnelles en 0 ne sont pas linéaires en v .

1.2 Le cas de la dimension finie

On passe maintenant au cas de la dimension finie. Plus précisément, on suppose dans cette partie que E est de dimension finie : $E = \mathbb{R}^n$.

1.2.1 Dérivées partielles

Définition 1.8. Soit U un ouvert de \mathbb{R}^n , $a \in U$ et $f : U \rightarrow F$. On dit que f admet au point a une dérivée partielle par rapport à la variable x_j si la fonction suivante

$$t \mapsto f(a_1, a_2, \dots, a_{j-1}, t, a_{j+1}, \dots, a_n)$$

est dérivable en $t = a_j$. Sa dérivée partielle par rapport à la variable x_j est la valeur de cette dérivée, et on la note par $\frac{\partial f}{\partial x_j}(a)$.

Ainsi, cette dérivée partielle est également donnée par la limite suivante (si elle existe, sinon la fonction f n'admet pas de dérivée partielle) :

$$\frac{\partial f}{\partial x_j}(a) = \lim_{\varepsilon \rightarrow 0} \frac{f(a + \varepsilon e_j) - f(a)}{\varepsilon}$$

où on a noté par e_j le j -ème élément de la base canonique de \mathbb{R}^n : toutes les composantes de e_j sont nulles sauf la j -ème qui est égale à 1.

La différentielle peut s'exprimer en fonction des dérivées partielles de la manière suivante.

Proposition 1.9. Soit U un ouvert de \mathbb{R}^n , $a \in U$ et $f : U \rightarrow F$ une fonction différentiable en a . Alors

$$Df(a)(h) = \sum_{j=1}^n h_j \frac{\partial f}{\partial x_j}(a).$$

On peut se demander quelle est la relation entre dérivées partielles et dérivées directionnelles et si l'existence des unes implique l'existence des autres. On remarque d'abord que les dérivées directionnelles concernent toutes les directions, et pas seulement celles des axes coordonnés. On pourrait donc penser que l'existence de toutes les dérivées directionnelles implique celles de toutes les dérivées partielles, mais ce n'est pas le cas à cause d'une subtilité : les dérivées directionnelles ne sont pas définies comme une limite pour $\varepsilon \rightarrow 0$, mais seulement d'un côté ($\varepsilon \rightarrow 0^+$) ; elles correspondent donc à une dérivée droite ou gauche seulement ; il serait donc possible que la restriction d'une fonction aux axes coordonnés passant par un point admette des dérivées droite et gauche mais différentes, et donc pas de dérivée tout court. On peut considérer les deux exemples suivants.

Exemples.

- L'existence des dérivées directionnelles suivant toute direction n'entraîne pas forcément celle des dérivées partielles. La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ donnée par $f(x) := \|x\|$ (en utilisant, par exemple, la norme euclidienne) admet des dérivées directionnelles en $a = 0$ données par $\frac{\partial f}{\partial v}(0) = \|v\|$ mais chaque fonction $t \mapsto f(a_1, a_2, \dots, a_{j-1}, t, a_{j+1}, \dots, a_n)$ coïncide avec $|t|$ et n'est donc pas dérivable en $t = 0 = a_j$.
- L'existence des dérivées partielles n'entraîne pas non plus celle des dérivées directionnelles, comme on peut voir dans l'exemple suivant d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$.

$$f(x, y) = \begin{cases} 0 & \text{si } xy = 0 \\ 1 & \text{si } xy \neq 0 \end{cases}$$

Comme la restriction de f à chaque axe coordonné passant par l'origine $a = (0, 0)$ est la fonction nulle, les dérivées partielles existent et sont nulles. D'autre part, si on prend un vecteur v qui n'est pas orienté comme les axes (par exemple $v = (1, 1)$) on a $f(a + \varepsilon v) = 1$ et $f(a) = 0$, ce qui entraîne que la limite définissant $\frac{\partial f}{\partial v}(0)$ vaut $+\infty$ et la fonction n'admet donc pas de dérivées directionnelles sauf pour certains v .

1.2.2 Matrice jacobienne et gradient

Supposons maintenant que F est lui aussi de dimension finie : $F = \mathbb{R}^m$. Une fonction f à valeurs dans F admet m composantes

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$$

Comme la limite dans \mathbb{R}^m se fait composante par composante et que les dérivées partielles sont définies via une limite, nous avons que f admet des dérivées partielles ssi chaque composante de f admet des dérivées partielles et la dérivée partielle de f s'obtient en prenant les dérivées partielles des composantes.

La différentielle est une application linéaire de \mathbb{R}^n dans \mathbb{R}^m . Elle s'identifie donc à une matrice. On peut exprimer cette matrice en fonction des dérivées partielles des composantes de f .

Proposition 1.10. *Soit U un ouvert de \mathbb{R}^n , $a \in U$ et $f : U \rightarrow \mathbb{R}^m$ une fonction différentiable en a . La matrice de $Df(a)$ dans les bases canoniques de \mathbb{R}^n et \mathbb{R}^m est donnée par la matrice suivante*

$$M_f(a) = \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \dots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

On appelle cette matrice la matrice jacobienne en a .

Une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ peut aussi être vue comme une fonction définie de \mathbb{C} dans \mathbb{C} . Nous avons alors deux notions de différentiabilité. D'une part la notion de différentiabilité sur \mathbb{R}^2 vu comme un \mathbb{R} -espace vectoriel. Et d'autre part la notion de différentiabilité sur \mathbb{C} vu comme un \mathbb{C} -espace vectoriel. Ces deux notions sont-elles les mêmes ? La réponse est non. Plus précisément, la \mathbb{C} différentiabilité implique la \mathbb{R} différentiabilité mais la réciproque est fautive. Cela vient du fait qu'une application linéaire sur \mathbb{C} est aussi linéaire sur \mathbb{R}^2 mais une application linéaire sur \mathbb{R}^2 ne l'est pas forcément sur \mathbb{C} .

Définition 1.11. *Soit U un ouvert de \mathbb{R}^n , $a \in U$ et $f : U \rightarrow \mathbb{R}$ une fonction différentiable en a . On appelle gradient de f en a le vecteur ligne*

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right).$$

Le gradient coïncide avec la matrice jacobienne. On peut facilement voir que le gradient est la direction où f augmente le plus vite.

1.2.3 Continuité des dérivées partielles et différentiabilité

Le critère le plus important pour la différentiabilité d'une fonction est celui de la continuité des dérivées partielles.

Théorème 1.12. *Soit U un ouvert de \mathbb{R}^n , $a \in U$ et $f : U \rightarrow \mathbb{R}^m$. On suppose que les dérivées partielles de f existent dans un voisinage de a et sont continues en a . Alors f est différentiable en a .*

Si la continuité des dérivées partielles est une condition suffisante de différentiabilité, ce n'est pas une condition nécessaire (seule l'existence des dérivées partielles est une condition nécessaire). En pratique, lorsqu'on veut décider de la différentiabilité d'une fonction concrète (qui a en général des points de singularité) on peut procéder de la manière suivante :

1. On étudie la continuité de la fonction. Si la fonction n'est pas continue elle n'est pas différentiable.
2. Si la fonction est continue, on étudie l'existence des dérivées partielles. Si au moins une des dérivées partielles n'existe pas, la fonction n'est pas différentiable.
3. Si la fonction est continue et toutes les dérivées partielles existent, on étudie la continuité des dérivées partielles. Si toutes les dérivées partielles sont continues alors la fonction est différentiable.
4. Si la fonction est continue et toutes les dérivées partielles existent mais certaines sont discontinues, il ne reste plus qu'à vérifier la définition de la différentiabilité. Mais l'application linéaire et continue L de la définition est connue (sa matrice est la matrice des dérivées partielles, la matrice jacobienne) donc la vérification de la définition est maintenant aisée.

Exemple. La fonction

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f(x, y) = \begin{cases} x^2 \sin \frac{1}{x^2+y^2} & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

est continue en $(0, 0)$, les dérivées partielles existent partout mais sont discontinues en $(0, 0)$. La fonction est différentiable en $(0, 0)$ de différentielle nulle.

Cela nous donne l'occasion de discuter la notion de fonction C^1 et plus généralement C^k .

Définition 1.13. — Une fonction $f : U$ ouvert de E espace normé à valeurs dans F espace normé est dite de classe C^1 si elle est différentiable et si sa différentielle $Df : U \rightarrow \mathcal{L}(E, F)$ est continue.

— Une fonction $f : U$ ouvert de E espace normé à valeurs dans F espace normé est dite de classe C^k , $k \geq 2$ si elle est différentiable et si sa différentielle $Df : U \rightarrow \mathcal{L}(E, F)$ est de classe C^{k-1} . Elle est dite de classe C^∞ si elle est de classe C^k pour tout k .

1.3 Inégalité des accroissements finis

Commençons par définir la dérivabilité.

Définition 1.14. Soit I un intervalle de \mathbb{R} , E un espace normé et $f : I \rightarrow E$. On dit que f est dérivable en un point t_0 de I si la limite

$$f'(t_0) = \lim_{t \rightarrow t_0} \frac{f(t) - f(t_0)}{t - t_0}$$

existe.

Pour une fonction d'une variable réelle à valeurs dans un espace normé, les notions de dérivabilité et différentiabilité coïncident.

Nous rappelons d'abord le théorème des accroissements finis pour des fonctions de \mathbb{R} dans \mathbb{R} .

Théorème 1.15. Si $f : [a, b] \rightarrow \mathbb{R}$ est une fonction continue, dérivable dans $]a, b[$, il existe un point $\xi \in]a, b[$ tel que $f'(\xi) = (f(b) - f(a))/(b - a)$.

On voudrait étendre ce théorème au cas multidimensionnel. Première mauvaise surprise : il est faux même pour des fonctions d'une variable, lorsqu'elles sont à valeur dans \mathbb{R}^m , même pour $m = 2$. On peut considérer l'exemple suivant

$$[a, b] = [0, 2\pi], \quad f(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

En effet $f(b) - f(a) = 0$, mais $f'(t)$ ne s'annule jamais comme vecteur (ses composantes s'annulent, mais pas en un même point).

Lemme 1.16. Soit I un intervalle de \mathbb{R} , E un espace normé, $f : I \rightarrow E$ et $t_0 \in I$. La fonction f est dérivable en t_0 si et seulement si elle est différentiable en t_0 . De plus, la différentielle en t_0 est l'application linéaire et continue de multiplication par la dérivée $f'(t_0)$ et nous avons

$$\|Df(t_0)\|_{\mathcal{L}(\mathbb{R}, E)} = \|f'(t_0)\|_E.$$

Voici l'inégalité des accroissements finis dans un cas particulier.

Lemme 1.17. Soit $f : [a, b] \rightarrow E$ dérivable. On suppose qu'il existe C tel que $\|f'(t)\|_E \leq C$ pour tout $t \in [a, b]$. Alors $\|f(b) - f(a)\|_E \leq C|b - a|$.

Enfin le cas général :

Théorème 1.18. Soit $f : \Omega \rightarrow F$ où Ω est un ouvert de E et E, F sont des espaces normés. Soient $a, b \in \Omega$ tels que le segment $[a, b] = \{ta + (1 - t)b ; t \in [0, 1]\}$ soit inclus dans Ω . On suppose que f est différentiable en tout point de $[a, b]$ et que la norme de sa différentielle en tout point du segment est bornée par une constante indépendante du point. Alors nous avons l'inégalité suivante

$$\|f(a) - f(b)\|_F \leq \|a - b\|_E \sup_{x \in [a, b]} \|Df(x)\|_{\mathcal{L}(E, F)}.$$

Corollaire 1.19. Soit $f : \Omega \rightarrow F$ où Ω est un ouvert convexe de E et E, F sont des espaces normés. On suppose que f est différentiable sur Ω et que sa différentielle est bornée. Alors f est Lipschitzienne.

Corollaire 1.20. Soit $f : \Omega \rightarrow F$ où Ω est un ouvert connexe de E et E, F sont des espaces normés. On suppose que f est différentiable sur Ω et que sa différentielle est nulle en tout point. Alors f est constante.

1.4 Dérivées d'ordre deux et supérieur

On considère une fonction $f : \Omega \rightarrow \mathbb{R}$, avec $\Omega \subset \mathbb{R}^n$ un ouvert (sinon, on peut la regarder composante par composante, si c'est une fonction à valeurs vectoriels). On a défini son gradient $\nabla f : \Omega \rightarrow \mathbb{R}^n$, si cette fonction est différentiable (l'existence des dérivées partielles suffit pour définir le gradient comme vecteur). On peut regarder la différentielle du gradient en un point $\bar{x} \in \Omega$, c'est une matrice carrée $n \times n$ appelée Hessienne, et notée $D^2 f(\bar{x})$. Ses composantes sont données par les dérivées secondes

$$\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) (\bar{x}).$$

Un point important est le fait que cette matrice est, sous une simple hypothèse de continuité, symétrique :

Théorème 1.21 (Théorème de Schwarz). Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $\bar{x} \in \Omega$, et $f : \Omega \rightarrow \mathbb{R}$ une fonction C^1 dans un voisinage de \bar{x} , telles que ses dérivées partielles admettent aussi des dérivées partielles par rapport à toutes les variables, et que ces dérivées secondes sont continues en \bar{x} . Alors on a

$$\frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) (\bar{x}) = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) (\bar{x}).$$

Ce théorème se démontre, entre autre, par une double application du théorème des accroissements finis pour des fonctions réelles d'une variable réelle. On indiquera désormais par $\partial^2 f / (\partial x_i \partial x_j)$ les dérivées partielles secondes, sans se soucier de l'ordre de dérivation.

Les dérivées secondes servent aussi à donner un développement limité d'ordre deux aux fonctions de plusieurs variables. Nous rappelons d'abord ce qu'un DL2 dans le cas d'une variable : si f est C^{k-1} et qu'elle admet une dérivée k -ème en \bar{x} on a

$$f(\bar{x} + h) = \sum_{j=0}^k \frac{1}{j!} f^{(j)}(\bar{x}) h^j + o(h^k)$$

lorsque $h \rightarrow 0$. Si on se limite aux cas $k = 1, 2$ on a

$$f(\bar{x} + h) = f(\bar{x}) + f'(\bar{x})h + o(h), \quad f(\bar{x} + h) = f(\bar{x}) + f'(\bar{x})h + \frac{f''(\bar{x})}{2}h^2 + o(h^2).$$

La première formule se généralise dans le cas de plusieurs variables en

$$f(\bar{x} + h) = f(\bar{x}) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\bar{x}) h_j + o(|h|),$$

et ce n'est rien d'autre que la définition de différentielle et son identification avec le produit scalaire avec le gradient. La deuxième par contre donne le DL que l'on trouve dans cet énoncé (qu'on présente par simplicité dans le cas C^2 , alors que la différentiabilité du gradient suffirait...)

Théorème 1.22. *Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $\bar{x} \in \Omega$, et $f : \Omega \rightarrow \mathbb{R}$ une fonction C^2 . Alors on a*

$$f(\bar{x} + h) = f(\bar{x}) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\bar{x}) h_j + \frac{1}{2} \sum_{l,k=1}^n \frac{\partial^2 f}{\partial x_l \partial x_k}(\bar{x}) h_l h_k + o(|h|^2).$$

Cela peut également s'écrire sous la forme

$$f(\bar{x} + h) = f(\bar{x}) + \nabla f(\bar{x}) \cdot h + \frac{1}{2} D^2 f(\bar{x}) h \cdot h + o(|h|^2)$$

Le développement à l'ordre deux permet de mieux comprendre le rôle de points critiques (ceux où les gradients s'annule). Pour cela il faut rappeler la notion de matrice (semi)définie positive(négative).

Définition 1.23. *Une matrice carrée $n \times n$ symétrique A est dit définie positive si $Ah \cdot h > 0$ pour tout vecteur $h \in \mathbb{R}^n \setminus \{0\}$. Cela se produit si et seulement si toutes ses valeurs propres sont strictement positives.*

Une matrice carrée $n \times n$ symétrique A est dit semi-définie positive si $Ah \cdot h \geq 0$ pour tout vecteur $h \in \mathbb{R}^n$. Cela se produit si et seulement si toutes ses valeurs propres sont non-négatives.

Une matrice carrée $n \times n$ symétrique A est dit définie négative si $Ah \cdot h < 0$ pour tout vecteur $h \in \mathbb{R}^n \setminus \{0\}$. Cela se produit si et seulement si toutes ses valeurs propres sont strictement négatives.

Une matrice carrée $n \times n$ symétrique A est dit semi-définie négative si $Ah \cdot h \leq 0$ pour tout vecteur $h \in \mathbb{R}^n$. Cela se produit si et seulement si toutes ses valeurs propres sont non-positives.

Nous avons le résultat suivant.

Théorème 1.24. *Soit $\Omega \subset \mathbb{R}^n$ un ouvert. Si un point $\bar{x} \in \Omega$ est un minimum(maximum) local de f alors $\nabla f(\bar{x}) = 0$ et la matrice $D^2 f(\bar{x})$ est semi-définie positive(négative). Viceversa, si un point $\bar{x} \in \Omega$ est tel que $\nabla f(\bar{x}) = 0$ et la matrice $D^2 f(\bar{x})$ est définie positive(négative) alors il est un minimum(maximum) local.*

Si un point $\bar{x} \in \Omega$ est tel que $\nabla f(\bar{x}) = 0$ mais la matrice $D^2 f(\bar{x})$ admet tant des valeurs propres positives que négatives, alors le point \bar{x} est un point selle (ou col), c'est-à-dire un point qui minimise localement f dans une direction et maximise localement dans une autre.

Pour présenter les DL d'ordre supérieur à deux il faut introduire la notation des multi-index. Un multi-index α est un vecteur de \mathbb{N}^n ; on a donc $\alpha = (\alpha_1, \dots, \alpha_n)$ avec $\alpha_i \in \mathbb{N}$ pour tout i . Nous utilisons les notations suivantes

$$|\alpha| := \sum_i \alpha_i; \quad \frac{\partial^\alpha f}{\partial x^\alpha} := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}; \quad x^\alpha := x_1^{\alpha_1} \dots x_n^{\alpha_n}; \quad \alpha! := \alpha_1! \dots \alpha_n!$$

Le résultat concernant les DL à l'ordre m est alors le suivant.

Théorème 1.25. *Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $\bar{x} \in \Omega$, et $f : \Omega \rightarrow \mathbb{R}$ une fonction C^m . Alors on a*

$$f(\bar{x} + h) = \sum_{k=0}^m \sum_{\alpha: |\alpha|=k} \frac{1}{\alpha!} \frac{\partial^\alpha f}{\partial x^\alpha}(\bar{x}) h^\alpha + o(\|h\|^m).$$

1.5 Théorème d'inversion locale

On présente d'abord un résultat préliminaire. On rappelle d'abord la définition de homéomorphisme.

Définition 1.26. *Une fonction f est dite homéomorphisme de U dans V si f est bijective de U dans V , et f et f^{-1} sont des applications continues.*

On remarque que la définition de continuité, et donc celle de homéomorphisme, ne nécessite pas la structure d'espace vectoriel normé (des espaces topologiques suffiraient). Dans le cas des espaces vectoriels, les applications linéaires qui sont des homéomorphismes (et dont la réciproque serait donc aussi linéaire) jouent un rôle particulier.

Proposition 1.27. *$L \in \mathcal{L}(E, F)$ est un homéomorphisme si et seulement si*

- L est surjective;
- il existe une constante $c > 0$ telle que $c\|h\|_E \leq \|L(h)\|_F$ pour tout h .

On regarde maintenant la différentielle de l'inverse, qui est l'inverse de la différentielle. Plus précisément :

Théorème 1.28. *Soient E et F deux espaces normés, U un ouvert de E , V un ouvert de F et $f : U \rightarrow V$ un homéomorphisme. On suppose que f est différentiable en un point $a \in U$ et que sa différentielle $Df(a)$ est un homéomorphisme de E dans F . Alors f^{-1} est différentiable en $f(a)$ et*

$$D(f^{-1})(f(a)) = (Df(a))^{-1}.$$

On passe ensuite à la notion de difféomorphisme.

Définition 1.29. *Une fonction f est dite difféomorphisme de U dans V si U et V sont ouverts, si f est bijective de U dans V , f est C^1 et f^{-1} est C^1 .*

Montrons un lemme qui sera utilisé dans la preuve du théorème d'inversion locale. On écrit $\mathcal{L}(E)$ pour les applications linéaires continues de E dans E (donc $\mathcal{L}(E, E)$)

Lemme 1.30. *Soit E un espace de Banach. L'ensemble $X = \{T \in \mathcal{L}(E) ; T \text{ inversible}\}$ est un ouvert de $\mathcal{L}(E)$ et l'application $X \ni T \mapsto T^{-1} \in \mathcal{L}(E)$ est continue.*

Voici le théorème d'inversion locale.

Théorème 1.31 (inversion locale). Soient E un espace de Banach, $f : U \rightarrow E$ (où U est un ouvert de E) une fonction de classe C^1 et $a \in U$ tels que $Df(a)$ est un homéomorphisme. Alors il existe U' un ouvert qui contient a et V' un ouvert qui contient $f(a)$ tels que f est un difféomorphisme de U' dans V' .

Nous avons aussi une version globale du théorème d'inversion locale.

Théorème 1.32 (inversion globale). Soient E un espace de Banach, $f : U$ ouvert de E à valeurs dans E une fonction de classe C^1 . Si f est injective et $Df(x)$ est un homéomorphisme pour tout $x \in U$ alors $f(U)$ est un ouvert et f est un difféomorphisme de U dans $f(U)$.

1.6 Théorème des fonctions implicites

Le théorème des fonctions implicites permet de résoudre une équation du type $f(x, y) = 0$ en exprimant une des variables en fonction des autres. Par exemple $y = \varphi(x)$ où φ est une fonction implicite. On sait que φ existe mais on ne la connaît pas explicitement, d'où la terminologie de fonction implicite. On peut montrer ainsi que les zéros d'une fonction de \mathbb{R}^2 se trouvent sur une courbe ; on discutera cela plus en détail quand on parlera de courbes plus tard.

Définissons d'abord la notion de différentielle partielle, qui est similaire à la notion de dérivée partielle.

Définition 1.33. Soit $f : U$ ouvert de $E \times F$ à valeurs dans G où E, F et G sont des espaces normés. On dit que $f = f(x, y)$ admet une différentielle partielle par rapport à x au point $(a, b) \in U$ si l'application $x \mapsto f(x, b)$ est différentiable en a et on note

$$D_x f(a, b) = D(x \mapsto f(x, b))(a).$$

On définit de même la différentielle partielle par rapport à y et on note

$$D_y f(a, b) = D(y \mapsto f(a, y))(b).$$

Comme dans le cas des dérivées partielles, on peut exprimer la différentielle en fonction des différentielles partielles.

Lemme 1.34. Soit $f : U$ ouvert de $E \times F$ à valeurs dans G où E, F et G sont des espaces normés. Si f est différentiable en (a, b) alors elle admet des différentielles partielles en (a, b) et on a

$$Df(a, b)(h, k) = D_x f(a, b)h + D_y f(a, b)k.$$

Voici le théorème des fonctions implicites.

Théorème 1.35 (fonctions implicites). Soient E, F des espaces de Banach, U un ouvert de $E \times F$, $f : U \rightarrow F$ une fonction de classe C^1 . Soit $(a, b) \in U$ tel que $f(a, b) = 0$ et $D_y f(a, b)$ est un homéomorphisme de $\mathcal{L}(F)$. Alors il existe un ouvert W qui contient (a, b) , un ouvert V qui contient a et une fonction $\varphi : V \rightarrow F$ de classe C^1 tels qu'on a l'équivalence suivante :

$$(x, y) \in W \text{ et } f(x, y) = 0 \quad \Leftrightarrow \quad x \in V \text{ et } y = \varphi(x).$$

Exemples.

a) Considérons le système

$$\begin{cases} 4xy + 2xz + y + 4y^2 = 0 \\ x^3y + xz + 4z - z^2 = 0 \end{cases}$$

au voisinage du point $(0, 0, 0)$. Le théorème des fonctions implicites s'applique et nous permet d'exprimer y et z en fonction de x . Mais il ne s'applique pas pour exprimer x et y en fonction de z , ni x et z en fonction de y . Par ailleurs, il ne peut pas s'appliquer pour exprimer par exemple x en fonction de y et z car dans le théorèmes des fonctions implicites le nombre de variables qui s'expriment en fonction des autres est toujours égal au nombre d'équations.

- b) Considérons l'équation $2xy - z + 2xz^3 = 5$ au voisinage du point $(1, 2, 1)$. On peut exprimer z en fonction de x et y et on peut calculer les dérivées de la fonction implicite en $(1, 2)$ à n'importe quel ordre.

2 Courbes et surfaces paramétrées

2.1 Définitions et représentation implicite

Définition 2.1. — Une courbe paramétrée de \mathbb{R}^n est une application continue $\gamma : I \rightarrow \mathbb{R}^n$ où I est un intervalle de \mathbb{R} .

- Le support de la courbe est $\gamma(I)$.
- Une courbe est dite de classe C^k si l'application γ est de classe C^k .
- Une courbe est dite simple si l'application γ est injective.
- Deux courbes paramétrées $\gamma : I \rightarrow \mathbb{R}^n$ et $\tilde{\gamma} : \tilde{I} \rightarrow \mathbb{R}^n$ sont dites équivalentes s'il existe un homéomorphisme $\varphi : I \rightarrow \tilde{I}$ tel que $\tilde{\gamma} \circ \varphi = \gamma$. L'application φ est dite changement de paramètre, ou changement de paramétrage. Si φ est croissante, le sens de parcours de la courbe est conservé ; dans le cas contraire il est inversé. Si γ et $\tilde{\gamma}$ sont C^1 et φ est un difféomorphisme, on dit qu'elles sont C^1 -équivalentes. Si elles sont C^k et que φ est un difféomorphisme C^k , on dit qu'elles sont C^k -équivalentes.
- Un point $\gamma(t_0)$ d'une courbe γ est dit régulier si $\gamma'(t_0) \neq 0$. Une courbe est dite régulière si tous ses points sont réguliers.
- Si un point $\gamma(t_0)$ d'une courbe γ est régulier, la direction du vecteur $\gamma'(t_0)$ est dite direction tangente à γ en $\gamma(t_0)$.
- Un point $\gamma(t_0)$ d'une courbe $\gamma \in C^2$ est dit bi-régulier si $\gamma'(t_0) \neq 0$, $\gamma''(t_0) \neq 0$, et ces deux vecteurs ne sont pas colinéaires.

On pourrait avoir envie de considérer comme une courbe certains sous-ensembles de \mathbb{R}^n décrits de manière implicite comme zéros d'une fonction. Par exemple, l'ensemble

$$C := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 - 1 = 0\}$$

est clairement le cercle unité autour de l'origine, qui peut être paramétré par $\gamma(t) = (\cos t, \sin t)$ pour $t \in [0, 2\pi]$ (ou pour $t \in \mathbb{R}$, ou pour t dans tout intervalle de longueur au moins 2π). Cela se base pourtant sur une intuition qui nous a fait deviner la formule de γ . Le théorème des fonctions implicites nous permet de traiter le cas général, au moins localement.

Proposition 2.2. Si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est C^1 et $\nabla f \neq 0$ sur l'ensemble $C = \{x \in \mathbb{R}^2 : f(x) = 0\}$, alors localement cet ensemble peut être décrit comme graphe d'une fonction $\phi \in C^1$, c'est-à-dire comme support d'une courbe de la forme $t \mapsto \gamma(t) = (t, \phi(t))$ ou $t \mapsto \gamma(t) = (\phi(t), t)$.

On observe en passant que les courbes obtenues comme graphe sont toujours régulières, parce qu'une des composantes de γ' vaut 1.

En dimension supérieure on a :

Proposition 2.3. Si $f : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$ est C^1 et le rang de Df vaut $n - 1$ partout sur l'ensemble $C = \{x \in \mathbb{R}^n : f(x) = 0\}$, alors localement cet ensemble peut être décrit comme graphe d'une fonction d'une variable $\phi \in C^1$, c'est-à-dire, dans un système de coordonnées opportun, comme support d'une courbe de la forme $t \mapsto \gamma(t) = (t, \phi_1(t), \dots, \phi_{n-1}(t))$.

Dans le cas d'une courbe définie implicitement par $f = 0$, la direction tangente en un point x_0 où $f(x_0) = 0$ et le rang de $Df(x_0)$ vaut $n - 1$ est donnée par l'orthogonal à tous les vecteurs $\nabla f_i(x_0)$ pour $i = 1, \dots, n - 1$.

On peut faire de même pour les surfaces.

Définition 2.4. — Une surface paramétrée de \mathbb{R}^n est une application continue $X : \Omega \rightarrow \mathbb{R}^n$ où Ω est un ouvert connexe de \mathbb{R}^2 .

- Le support de la surface est $X(\Omega)$
- Une surface est dite de classe C^k si l'application X est de classe C^k .
- Une surface est dite simple si l'application X est injective.
- Deux surfaces paramétrées $X : \Omega \rightarrow \mathbb{R}^n$ et $\tilde{X} : \tilde{\Omega} \rightarrow \mathbb{R}^n$ sont dites équivalentes s'il existe un homéomorphisme $\varphi : \Omega \rightarrow \tilde{\Omega}$ tel que $\tilde{X} \circ \varphi = X$. L'application φ est dite changement de paramètre, ou changement de paramétrage. Si X et \tilde{X} sont C^1 et φ est un difféomorphisme, on dit qu'elles sont C^1 -équivalentes. Si elles sont C^k et que φ est un difféomorphisme C^k , on dit qu'elles sont C^k -équivalentes.
- Un point $X(u_0, v_0)$ d'une surface est dit régulier si $\frac{\partial X}{\partial u}(u_0, v_0)$ et $\frac{\partial X}{\partial v}(u_0, v_0)$ forment une famille libre. Une surface est dite régulière si tous ses points sont réguliers.
- Si un point $X(u_0, v_0)$ est régulier, l'espace vectoriel de dimension deux engendré par les vecteurs $\frac{\partial X}{\partial u}(u_0, v_0)$ et $\frac{\partial X}{\partial v}(u_0, v_0)$ est dit plan tangent à X en ce point.

Dans le cas des surfaces aussi, on pourrait avoir envie de considérer des sous-ensembles de \mathbb{R}^n décrits de manière implicite. Par exemple, l'ensemble

$$C := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 - 1 = 0\}$$

est clairement la sphère unité autour de l'origine, qui peut être paramétré par

$$X(u, v) = (\cos u \cos v, \sin u \cos v, \sin v)$$

pour (u, v) dans tout ouvert contenant $[0, 2\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$. Le théorème des fonctions implicites peut également être utilisé localement. On présentera les résultats seulement dans le cas des surfaces dans l'espace de dimension trois.

Proposition 2.5. Si $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ est C^1 et $\nabla f \neq 0$ sur l'ensemble $C = \{x \in \mathbb{R}^3 : f(x) = 0\}$, alors localement cet ensemble peut être décrit comme graphe d'une fonction $\phi \in C^1$, c'est-à-dire, dans un système de coordonnées opportun, comme support d'une surface de la forme $(u, v) \mapsto X(u, v) = (u, v, \phi(u, v))$.

On observe aussi que les surfaces paramétrées comme graphe sont toujours régulières.

Dans le cas d'une surface définie implicitement par $f = 0$, le plan tangent en un point x_0 où $f(x_0) = 0$ et $\nabla f(x_0) \neq 0$ est donné par l'orthogonal à $\nabla f(x_0)$.

2.2 Courbes paramétrées : longueur, courbure et torsion

2.2.1 Longueur et paramétrisation par longueur d'arc

Définition 2.6. La longueur d'une courbe paramétrée $\gamma : [a, b] \rightarrow \mathbb{R}^n$ est la borne supérieure des longueurs de toutes les lignes polygonales dont les sommets sont pris dans l'ordre sur la courbe :

$$\text{long}(\gamma) = \sup_{a=t_0 < t_1 < \dots < t_n = b} \sum_{j=1}^n \|\gamma(t_j) - \gamma(t_{j-1})\|,$$

le sup étant pris sur N et sur les familles de points (t_0, t_1, \dots, t_N) .

Proposition 2.7. Soit $\gamma : [a, b] \rightarrow \mathbb{R}^n$ une courbe de classe C^1 . Nous avons alors

$$\text{long}(\gamma) = \int_a^b \|\gamma'(t)\| dt.$$

Remarque. Deux courbes équivalentes ont la même longueur.

Définition 2.8. Une courbe est dite paramétrée par la longueur d'arc, ou par l'abscisse curviligne, si $\|\gamma'(t)\| = 1$ pour tout t .

Remarque. La notion de longueur d'une courbe, ainsi que celle de courbe paramétrée par la longueur d'arc, dépendent de la norme choisie. En l'absence d'indication contraire, on sous-entend en général qu'il s'agit de la norme euclidienne.

Proposition 2.9. Toute courbe régulière admet un paramétrage par la longueur d'arc. Ce paramétrage est unique à une translation près et au sens de parcours près. De plus, dans le cas de la norme euclidienne, si la courbe originale est C^k son paramétrage par la longueur d'arc l'est aussi.

Proposition 2.10. Dans un point régulier $\gamma(t_0)$, la direction de la tangente à la courbe est donnée par $\gamma'(t_0)$. Si, de plus, γ est paramétrée par la longueur d'arc, alors le vecteur $\gamma''(t_0)$ est normal à la courbe.

2.2.2 Allure d'une courbe plane

Dans toute cette partie nous considérons des courbes à valeur dans \mathbb{R}^2 muni de la norme euclidienne.

Définition 2.11. Soit γ une courbe C^2 paramétrée par la longueur d'arc. On appelle courbure la quantité

$$\kappa(t) = \|\gamma''(t)\|$$

et centre de courbure le point

$$P(t) = \gamma(t) + \frac{\gamma''(t)}{\|\gamma''(t)\|^2}$$

Proposition 2.12. Soit γ une courbe régulière C^2 (pas nécessairement paramétrée par la longueur d'arc). Nous avons la formule suivante pour la courbure :

$$\kappa(t) = \frac{|\gamma'(t) \wedge \gamma''(t)|}{\|\gamma'(t)\|^3}$$

où on a utilisé la notation $x \wedge y = x_1y_2 - x_2y_1$.

La courbure est une mesure quantitative du caractère 'plus ou moins courbé'. La courbure d'une droite est nulle. Celle d'un cercle de rayon R est $\frac{1}{R}$. Intuitivement, un bout d'un cercle de rayon très grand semble être presque plat ; il est donc normal que sa courbure soit petite. Nous avons aussi la réciproque :

Proposition 2.13. a) Toute courbe régulière de classe C^2 de courbure nulle est un bout de droite.

b) Toute courbe régulière de classe C^3 de courbure constante et strictement positive est un bout de cercle.

Définition 2.14. Le cercle osculateur est le cercle dont le centre est le centre de courbure et le rayon est l'inverse de la courbure : $C(P(t), \frac{1}{\kappa(t)})$.

2.2.3 Remarques sur les courbes gauches

Une courbe gauche est une courbe de \mathbb{R}^3 qui n'est pas plane. La tangente à la courbe est encore dirigée par γ' . Soit γ paramétrée par la longueur d'arc. Alors γ'' est encore orthogonale à γ' , donc à la courbe. On appelle plan osculateur le plan engendré par $\gamma'(t)$ et $\gamma''(t)$ et qui passe par $\gamma(t)$. On peut définir comme en dimension deux la courbure et le cercle osculateur qui sera tracé dans le plan osculateur. Une quantité spécifique à la dimension trois est la torsion qui est définie par

$$\theta = -\frac{\det(\gamma', \gamma'', \gamma''')}{\|\gamma' \wedge \gamma''\|^2}$$

La torsion d'une courbe mesure la manière dont la courbe se tord pour sortir de son plan osculateur. On peut montrer

Proposition 2.15. *Soit γ une courbe C^3 dans \mathbb{R}^3 , birégulière en tout point. Alors sa torsion est identiquement nulle si et seulement si la courbe est plane.*

2.3 Géodésiques sur la sphère et sur d'autres surfaces

Définition 2.16. *Étant donné un sous-ensemble fermé et connexe $M \subset \mathbb{R}^n$ on dit qu'une courbe C^1 $\gamma : [0, T] \rightarrow M$ est une géodésique de M si elle satisfait la propriété suivante :*

$$\text{long}(\gamma) = \min\{\text{long}(\tilde{\gamma}) : \tilde{\gamma} : [0, T] \rightarrow M, \tilde{\gamma} \in C^1, \tilde{\gamma}(0) = \gamma(0), \tilde{\gamma}(T) = \gamma(T)\}.$$

On se propose de trouver les courbes qui sont géodésiques dans le cas de la sphère $M = \mathbb{S}^2 \subset \mathbb{R}^3 = \{x \in \mathbb{R}^3, \|x\| = 1\}$ (la norme étant toujours la norme euclidienne). Il s'agit en fait des arcs de grand cercle, qui sont de la forme $\gamma(t) = v \cos t + w \sin t$ pour tout choix de v, w vecteurs unités orthogonaux : $\|v\| = \|w\| = 1, v \cdot w = 0$, à condition que T soit suffisamment petit.

On a besoin des lemmes suivants.

Lemme 2.17. *Il existe une constante C telle que l'inégalité suivante (dite de Poincaré) est satisfaite par toute fonction $\eta : [0, T] \rightarrow \mathbb{R}^d$ de classe C^1 avec $\eta(0) = \eta(T) = 0$:*

$$\int_0^T \|\eta(t)\|^2 dt \leq CT^2 \int_0^T \|\eta'(t)\|^2 dt.$$

Il est facile de montrer l'inégalité avec $C = 1$, de l'améliorer avec $C = 1/4$, mais la constante optimale (qu'on peut trouver en regardant les séries de Fourier) est $C = \pi^{-2}$.

Lemme 2.18. *Si $\gamma : [0, T] \rightarrow \mathbb{R}^3$ est de la forme $\gamma(t) = v \cos t + w \sin t$ avec $\|v\| = \|w\| = 1, v \cdot w = 0$, et $\tilde{\gamma} : [0, T] \rightarrow \mathbb{S}^2$ satisfait $\tilde{\gamma} \in C^1, \tilde{\gamma}(0) = \gamma(0), \tilde{\gamma}(T) = \gamma(T)$, et T est suffisamment petit ($CT^2 \leq 1$ si C est la constante du lemme précédent) alors on a*

$$\int_0^T \|\gamma'(t)\|^2 dt \leq \int_0^T \|\tilde{\gamma}'(t)\|^2 dt.$$

En choisissant $\tilde{\gamma}$ paramétrée à vitesse constante on obtient enfin :

Proposition 2.19. *Toute courbe $\gamma : [0, T] \rightarrow \mathbb{R}^3$ de la forme $\gamma(t) = v \cos t + w \sin t$ avec $\|v\| = \|w\| = 1, v \cdot w = 0$, est une géodésique de \mathbb{S}^2 si T est suffisamment petit.*

On peut se demander qu'en est-il des géodésiques sur d'autres surfaces. On considérera une surface S décrite de manière implicite comme $S = \{x \in \mathbb{R}^3 : f(x) = 0\}$ pour une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ régulière (au moins C^1 , mais on aura besoin jusqu'à C^3 pour certains énoncés) et telle que $\nabla f \neq 0$ sur S .

Le point clé de l'analyse d'optimalité des grands cercles étaient l'équation différentielle qu'ils satisfont, $\gamma'' = -\gamma$. Comme dans la sphère le vecteur normal au point $x \in \mathbb{S}^2$ est le vecteur unité x même, la généralisation qui nous intéresse concerne les courbes telles que l'accélération γ'' est normale à S en tout point. On a d'abord l'énoncé suivant.

Lemme 2.20. *Si une courbe $\gamma : [0, T] \rightarrow S$ est telle que γ'' est normal à S en tout point, alors on a forcément*

$$\gamma''(t) = -\frac{D^2 f(\gamma(t))(\gamma'(t), \gamma'(t))}{\|\nabla f(\gamma(t))\|^2} \nabla f(\gamma(t)).$$

On considère maintenant l'EDO définie par la relation ci-dessus, et on a :

Proposition 2.21. *Pour tout $x_0 \in S$ et $w \in \mathbb{R}^3$ tel que $w \cdot \nabla f(x_0) = 0$ le problème de Cauchy ci-dessous admet une solution locale unique*

$$\begin{cases} \gamma''(t) = -\frac{D^2 f(\gamma(t))(\gamma'(t), \gamma'(t))}{\|\nabla f(\gamma(t))\|^2} \nabla f(\gamma(t)), \\ \gamma(0) = x_0, \\ \gamma'(0) = w. \end{cases}$$

De plus, cette solution satisfait $\gamma(t) \in S$ et $\|\gamma'(t)\| = \|w\|$ pour tout t . Enfin, toute courbe $\gamma : [0, T] \rightarrow S$ solution de cete équation différentielle est une géodésique si T est suffisamment petit.

2.4 Première et seconde formes fondamentales

Considérons une surface C^1 régulière $S = X(\Omega) \subset \mathbb{R}^3$. En un point $P = X(u_0, v_0)$ l'espace tangent $\text{Tan}_P(S)$ est l'espace de dimension deux engendré par les vecteurs $\mathbf{u} := X_u(u_0, v_0)$ et $\mathbf{v} := X_v(u_0, v_0)$. Le vecteur normal \mathbf{n} à S en P est un vecteur unité orthogonal à \mathbf{u} et \mathbf{v} , donc on peut prendre $\mathbf{n} = \frac{\mathbf{u} \wedge \mathbf{v}}{\|\mathbf{u} \wedge \mathbf{v}\|}$. On définit deux formes bilinéaires sur $\text{Tan}_P(S)$.

La première, notée I, s'appelle *Première Forme fondamentale* et coïncide avec la restriction du produit scalaire Euclidien usuel à $\text{Tan}_P(S)$. Autrement dit, on a $I(w_1, w_2) := w_1 \cdot w_2$. Dans la base (\mathbf{u}, \mathbf{v}) cette forme bilinéaire s'écrit à l'aide de la matrice suivante :

$$I := \begin{pmatrix} \|\mathbf{u}\|^2 & \mathbf{u} \cdot \mathbf{v} \\ \mathbf{u} \cdot \mathbf{v} & \|\mathbf{v}\|^2 \end{pmatrix} = \begin{pmatrix} \|X_u\|^2 & X_u \cdot X_v \\ X_u \cdot X_v & \|X_v\|^2 \end{pmatrix}.$$

On note que cette matrice coïncide également avec la matrice $DX^T \cdot DX$. Cete matrice dépend évidemment de la paramétrisation, et si on écrit $X = Y \circ \phi$ (avec $X : \Omega \rightarrow \mathbb{R}^3$, $Y : \omega' \rightarrow \mathbb{R}^3$ et $\phi : \Omega' \rightarrow \Omega$) on trouve $I_X = (D\phi)^T \cdot I_Y \cdot D\phi$, où I_Y représente la première forme fondamentale écrite selon la base donnée par la paramétrisation Y (au point $\phi(u, v)$) et I_X celle écrite selon la base donnée par la paramétrisation X (au point (u, v)).

La première forme fondamentale est celle qui donne la déformation des longueurs entre les courbes dans Ω et celles sur S : si $\omega : [0, T] \rightarrow \Omega$ est une courbe, $\gamma := X \circ \omega$ est ue courbe à valeur dans S , et on a $\|\gamma'(t)\|^2 = \|DX(\omega(t))\dot{\omega}(t)\|^2 = I(\omega'(t), \omega(t))$, donc

$$\text{long}(\gamma) = \int_0^T \sqrt{I(\omega'(t), \omega(t))} dt.$$

La *Seconde Forme fondamentale*, notée II, est définie seulement lorsque X est C^2 . Elle représente l'écart dans la direction normale, à l'ordre deux, de la surface par rapport au plan tangent. Notamment,

pour tout $w \in \text{Tan}_P(S)$, la valeur de $\text{II}(w, w)$ est telle que l'on a $\mathbf{n} \cdot (Q - P) = \frac{1}{2}\text{II}(w, w) + o(\|w\|^2)$, où Q est le point de S tel que $Q - (P + w)$ est parallèle à \mathbf{n} . Si on prend $(u_0, v_0) = (0, 0)$ et $w = u\mathbf{u} + v\mathbf{v}$ on a

$$X(u, v) = P + w + \frac{1}{2}X_{uu}u^2 + X_{uv}uv + \frac{1}{2}X_{vv}v^2 + o(\|(u, v)\|^2),$$

ce qui nous donne, en prenant le produit scalaire avec \mathbf{n} et en utilisant $\mathbf{n} \cdot \mathbf{u} = \mathbf{n} \cdot \mathbf{v} = 0$,

$$\mathbf{n} \cdot (X(u, v) - P) = \frac{1}{2}X_{uu}u^2 + X_{uv}uv + \frac{1}{2}X_{vv}v^2 + o(\|X(u, v) - P\|^2),$$

ce qui nous permet d'écrire II dans la base (\mathbf{u}, \mathbf{v}) comme

$$\text{II} := \begin{pmatrix} X_{uu} \cdot \mathbf{n} & X_{uv} \cdot \mathbf{n} \\ X_{uv} \cdot \mathbf{n} & X_{vv} \cdot \mathbf{n} \end{pmatrix}.$$

Ici on a $\text{II} = D^2(X \cdot \mathbf{n})$. En cas de changement de paramétrisation $X = Y \circ \phi$, on utilise $D^2X = D^2Y \circ \phi(D\phi, D\phi) + DY \circ \phi \cdot D^2\phi$. en prenant le produit scalaire avec \mathbf{n} et en utilisant que DY est tangent, on trouve

$$\text{II}_X = \text{II}_Y(D\phi, D\phi) = (D\phi)^T \cdot \text{II}_Y \cdot D\phi.$$

Cela montre que, si les deux formes fondamentales dépendent en effet de la paramétrisation, on a

$$(\text{I}_X)^{-1}\text{II}_X = (D\phi)^T \cdot (\text{I}_Y)^{-1}\text{II}_Y \cdot D\phi,$$

et les deux matrices $(\text{I}_X)^{-1}\text{II}_X$ et $(\text{I}_Y)^{-1}\text{II}_Y$ sont conjuguées. Elles ont en particulier le même déterminant $\frac{\det \text{II}}{\det \text{I}}$, qui est un invariant (dit *courbure de Gauss*).

Dans le cas où S est exprimée comme un graph $X(u, v) = (u, v, \varphi(u, v))$ et que $\nabla\varphi(u_0, v_0) = 0$ (graphe qui est horizontal en (u_0, v_0)) le calcul des deux formes fondamentales est simple et on a

$$\text{I} = I, \quad \text{II} = D^2\varphi.$$

Dans le cas où S est exprimée comme $S = \{x : f(x) = 0\}$ pour une fonction $f \in C^2$ et telle que $\nabla f \neq 0$ sur S on a aussi l'identité entre formes bilinéaires

$$\text{II} = -\frac{D^2f}{\|\nabla f\|},$$

parce qu'on a, par exemple, $\nabla f(X) \cdot X_u u = -D^2f(X)(X_u, x_u)$. Attention : les deux formes bilinéaires pourraient être écrites dans deux bases différentes (d'habitude, quand on écrit D^2f comme matrice des dérivées secondes, on utilise la base canonique, pas celle engendrée par X_u et X_v).

La seconde forme fondamentale sert à déterminer si la surface se trouve d'un côté ou de l'autre du plan tangent : elle est du côté du vecteur \mathbf{n} si II est définie positive, du côté opposé si II est définie négative, et elle passe d'un côté à l'autre si II a deux valeurs propres de signes opposés (et donc si $\det \text{II} < 0$, ce qui est un invariant parce que le signe du déterminant de I est toujours positif).

Par exemple, pour $S = \{(x, y, z) : z^2 = x^2 + y^2 + 1\}$ la seconde forme fondamentale en $(0, 0, 1)$ est définie positive et S y est localement au-dessus du plan tangent, qui est horizontal. Pour la surface de rotation $S = \{(x, y, z) : z^2 + 1 = x^2 + y^2\}$, au contraire, la seconde forme fondamentale en $(1, 0, 0)$ n'est pas définie, et en effet la surface se trouve d'un côté en suivant $z = 0$ et de l'autre en suivant $y = 0$.

On considère maintenant une géodésique $\gamma : [0, T] \rightarrow S = \{f = 0\}$ avec $\gamma(t_0) = x_0$ et $\gamma'(t_0) = w$ et on regarde sa courbure. On prendra des géodésiques satisfaisant l'EDO introduite précédemment et on suppose $\|w\| = 1$ de manière à ce que la courbe soit paramétrée par longueur d'arc. On a

$$\kappa = \|\gamma''(t_0)\| = \frac{|D^2f(x_0)(w, w)|}{\|\nabla f(x_0)\|} = |\text{II}(w, w)|.$$

Cela montre le lien entre seconde forme fondamentale et courbures : les courbures des géodésiques passant par x_0 sont obtenues en évaluant la seconde forme fondamentale sur des vecteurs unités (donc des vecteurs où la première forme fondamentale vaut 1).

3 Éléments d'optimisation

3.1 Existence du minimum

On s'intéresse à des problèmes du type

$$\min\{f(x) : x \in E\},$$

où $f : E \rightarrow \mathbb{R}$ est une fonction continue et E est un sous-ensemble de \mathbb{R}^d . Dans la modélisation il est parfois utile de considérer des fonctions à valeurs dans $\mathbb{R} \cup \{+\infty\}$ (la valeur $-\infty$ par contre rend le problème de minimisation triviale) mais il faut comprendre que, dans les problèmes d'optimisation, les fonctions à minimiser ne peuvent jamais être à valeurs vectoriels, puisqu'il faut un ordre dans l'espace d'arrivée. Considérer des problèmes de maximisation au lieu de regarder la minimisation ne revient qu'à changer des signes.

On rappelle le théorème de Weierstrass.

Théorème 3.1. *Si f est continue et E est compact, alors il existe un point $x_0 \in E$ réalisant le minimum de f , c'est-à-dire $f(x_0) \leq f(x)$ pour tout $x \in E$.*

Souvent l'ensemble E n'est pas compact, mais on peut prouver l'existence lorsque la fonction f tend vers l'infini en l'infini ou, plus généralement, si sa limite (ou liminf) en l'infini est plus grande que des valeurs prises dans des points de E . L'énoncé le plus général qu'on peut prouver est le suivant.

Théorème 3.2. *Soit E un ensemble fermé mais non borné et $f : E \rightarrow \mathbb{R}$ une fonction continue.*

Soit $\ell_0 := \liminf_{\|x\| \rightarrow +\infty, x \in E} f(x)$. Alors

- *S'il existe $x_0 \in E$ tel que $f(x_0) < \ell_0$ alors f admet un minimum sur E .*
- *Si pour tout $x \in E$ on a $f(x) > \ell_0$ alors f n'admet pas de minimum sur E .*
- *Si pour tout $x \in E$ on a $f(x) \geq \ell_0$ et qu'il existe au moins un point $x_0 \in E$ tel que $f(x_0) = \ell_0$, alors f admet un minimum sur E .*

3.2 Conditions nécessaires d'optimalité

On rappelle cet énoncé bien connu :

Théorème 3.3. *Supposons que f est C^1 , que x_0 est un minimum local de f sur E , et que x_0 est dans l'intérieur de E (il existe un rayon $r > 0$ tel que $B(x_0, r) \subset E$). Alors $\nabla f(x_0) = 0$.*

On s'intéresse maintenant au cas où l'ensemble E est donné de manière implicite par des équations $g_i(x) = 0$. En général, cela définit une surface et aucun point de E ne serait à l'intérieur. On a le théorème suivant (*extrema liés* ou *multiplicateurs de Lagrange*) :

Théorème 3.4. *Supposons que $f : \mathbb{R}^d \rightarrow \mathbb{R}$ et $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ sont des fonctions C^1 , avec $k < n$, et posons $E = \{x \in \mathbb{R}^d : g(x) = 0\}$. Supposons que x_0 est un minimum local de f sur E , et que la matrice $Dg(x_0)$ est de rang k . Alors on peut trouver des nombres $\lambda_i \in \mathbb{R}$ tels que $\nabla f(x_0) = \sum_i \lambda_i \nabla g_i(x_0)$, où les fonctions g_i sont les composantes de g .*

La condition sur le rang de $Dg(x_0)$ est dite condition de *qualification des contraintes*. Ce théorème est à interpréter surtout comme une manière d'exclure l'optimalité de certains (en fait, la plupart) points : seuls les points où la condition $\nabla f(x_0) = \sum_i \lambda_i \nabla g_i(x_0)$ ou ceux où le rang de $Dg(x_0)$ est strictement plus petit que k ont le droit d'être des points de minimum.

Par exemple, pour la résolution d'un problème du type

$$\min\{f(x) : g_i(x) \leq 0\}$$

on obtient la recette suivante

- prouver d'abord l'existence d'un minimum (en prouvant par exemple que E est borné – il serait donc compact parce que fermé grâce aux inégalités larges dans la définition – ou en regardant la liminf de f en l'infini) ;
- faire une liste de tous les candidats à être des points de minimum :
 - tous les points au voisinage desquels f ou g ne seraient pas C^1 ;
 - tous les points à l'intérieur de E , où l'on a $g_i(x_0) < 0$ pour tout i et $\nabla f(x_0) = 0$;
 - pour chaque $J \subset \{1, \dots, k\}$, tous les points x_0 qui s'obtiennent en résolvant le système

$$\begin{cases} \nabla f(x_0) = \sum_{i \in J} \lambda_i \nabla g_i(x_0), \\ g_i(x_0) = 0 & \text{pour tout } i \in J \\ g_i(x_0) < 0 & \text{pour tout } i \notin J \end{cases}$$

où cela est un système avec $n + \#J$ inconnues et $n + \#J$ équations et devrait raisonnablement avoir peut de solutions ; ce système peut également s'écrire sous la forme

$$\nabla f(x_0) = \sum_{i=1}^k \lambda_i \nabla g_i(x_0), \quad \lambda_i g_i(x_0) = 0;$$

- tous les points où on ne peut pas appliquer le théorème des extrema liés parce que le rang de $Dg(x_0)$ n'est pas k (cela correspond à résoudre, pour tout J , un système avec $\#J$ équations pour imposer $g_i = 0$ et énormément d'équations pour imposer que tous les mineurs de la matrice Jacobienne s'annulent, ce qui n'a souvent pas de solutions).
- espérer que ces points ne soient pas trop nombreux et calculer la valeur de f en chacun ;
- prendre le ou les points qui donnent la valeur la plus petite.

On peut aussi donner une condition nécessaire d'optimalité d'ordre deux.

Théorème 3.5. *Supposons que $f : \mathbb{R}^d \rightarrow \mathbb{R}$ et $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ sont des fonctions C^2 , avec $k < n$, et posons $E = \{x \in \mathbb{R}^d : g(x) = 0\}$. Supposons que x_0 est un minimum local de f sur E , et que la matrice $Dg(x_0)$ est de rang k . On sait qu'on peut trouver des nombres $\lambda_i \in \mathbb{R}$ tels que $\nabla f(x_0) = \sum_i \lambda_i \nabla g_i(x_0)$.*

Le point x_0 doit alors satisfaire aussi

$$\left(D^2 f(x_0) - \sum_i \lambda_i D^2 g_i(x_0) \right) (v, v) \geq 0$$

pour tout vecteur v tel que $v \cdot \nabla g_i(x_0) = 0$ pour tout i .

Une application des multiplicateurs de Lagrange permet de prouver le résultat bien connu de diagonalisation des matrices symétriques :

Proposition 3.6. *Soit A une matrice $n \times n$ symétrique et x_1 une solution de*

$$\min \left\{ \frac{1}{2} Ax \cdot x : \|x\| = 1 \right\},$$

ainsi que, pour tout $k = 2, \dots, n$ x_k une solution de

$$\min \left\{ \frac{1}{2} Ax \cdot x : \|x\| = 1, x \cdot x_j = 0 \text{ pour tout } j < k \right\}.$$

Les vecteurs x_1, \dots, x_n forment alors une base orthonormée de \mathbb{R}^n et on a $Ax_k = \lambda_k x_k$ pour des nombres λ_k avec $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

On peut d'ailleurs vérifier la validité de la condition d'ordre deux sur les points solution des problèmes d'optimisation de cet exemple.

3.3 Fonctions convexes

Définition 3.7. Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite convexe si pour tout $x, y \in \mathbb{R}^n$ et tout $t \in [0, 1]$ on a

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

On dit que f est strictement convexe si la même inégalité est satisfaite au sens stricte dès que $x \neq y$ et $t \in]0, 1[$.

En dimension 1 on a les caractérisations suivantes.

Proposition 3.8. Si f est C^1 elle est convexe si et seulement si f' est une fonction croissante et si et seulement si on a l'inégalité suivante

$$f(y) \geq f(x) + f'(x)(y - x)$$

pour tout x, y .

Si f est C^2 alors elle est convexe si et seulement si $f'' \geq 0$.

En dimension supérieure cela devient

Proposition 3.9. Si f est C^1 elle est convexe si et seulement si ∇f satisfait l'inégalité

$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq 0$$

pour tout x, y et si et seulement si on a l'inégalité suivante

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

pour tout x, y .

Si f est C^2 alors elle est convexe si et seulement si $D^2 f \geq 0$ au sens des matrices symétriques.

On donne aussi la définition de fonction uniformément convexe ou elliptique.

Définition 3.10. Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite uniformément convexe ou elliptique s'il existe $\alpha > 0$ tel que $x \mapsto f(x) - \frac{\alpha}{2}\|x\|^2$ est une fonction convexe (on dit alors qu'elle est α -elliptique).

On obtient donc les caractérisations suivantes

Proposition 3.11. Si f est C^1 elle est α -elliptique si et seulement si ∇f satisfait l'inégalité

$$(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha\|x - y\|^2$$

pour tout x, y et si et seulement si on a l'inégalité suivante

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x) + \frac{\alpha}{2}\|x - y\|^2$$

pour tout x, y .

Si f est C^2 alors elle est α -elliptique si et seulement si $D^2 f \geq \alpha I$ au sens des matrices symétriques (donc si toutes ses valeurs propres sont supérieures ou égales à α).

On remarque que toute fonction elliptique est strictement convexe.

Pour la minimisation des fonctions convexes on a :

Proposition 3.12. Si f est une fonction convexe C^1 alors un point \bar{x} minimise f si et seulement si $\nabla f(\bar{x}) = 0$. Si f est une fonction strictement convexe alors le point de minimum, s'il existe, est unique. Si f est elliptique alors le point de minimum existe et est unique.

En effet, si la stricte convexité est suffisante pour l'unicité du minimiseur elle ne l'est pas pour l'existence (penser à $f(x) = e^x$). Par contre, l'ellipticité est suffisante pour l'existence parce que toute fonction elliptique est minorée par une parabole.

3.4 Algorithmes itératifs

On considère l'algorithme d'optimisation le plus simple, celui du gradient à pas fixe : étant donné un point x_0 , on définit une suite itérée en prenant $x_{k+1} = x_k - \tau \nabla f(x_k)$.

On a le théorème suivant.

Théorème 3.13. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction C^2 avec $\alpha I \leq D^2 f(x) \leq LI$ pour deux constants $L \geq \alpha > 0$. Supposons $\tau \in]0, \frac{2}{L}]$. Alors la suite définie par l'algorithme du gradient à pas fixe converge vers le seul minimiseur \bar{x} de f et on a*

$$\|x_k - \bar{x}\| \leq \|x_0 - \bar{x}\| \lambda^k$$

où le nombre λ est donné par $\lambda = \max\{1 - \tau\alpha, \tau L - 1\} < 1$.

La valeur de τ qui minimise la valeur de λ serait $\tau = \frac{2}{\alpha+L}$, qui donnerait $\lambda = \frac{L-\alpha}{L+\alpha}$.

Comme on sait que pour une fonction convexe la minimisation équivaut à annuler le gradient, on pourrait aussi considérer des algorithmes pour résoudre $F = 0$, en prenant $F = \nabla f$. Un des algorithmes les plus efficaces pour résoudre $F = 0$ est l'algorithme de Newton, qui consiste à chaque étape à trouver x_{k+1} en trouvant le point où le DL_1 de F autour de x_k s'annule, donc

$$x_{k+1} = x_k - (DF(x_k))^{-1} F(x_k).$$

Cet algorithme converge très vite mais à condition de démarrer en un point x_0 qui soit déjà suffisamment proche de \bar{x} solution de $F(\bar{x}) = 0$.

Théorème 3.14. *Soit $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction C^2 avec $F(\bar{x}) = 0$ et $DF(\bar{x})$ inversible. Alors il existe un rayon $R > 0$ tel que*

- $DF(x)$ est inversible pour tout $x \in B(\bar{x}, R)$;
- pour tout point $x \in B(\bar{x}, R)$ on a $x - (DF(x))^{-1} F(x) \in B(\bar{x}, R)$;
- pour tout point $x_0 \in B(\bar{x}, R)$ la suite donnée par $x_{k+1} = x_k - (DF(x_k))^{-1} F(x_k)$ converge vers \bar{x} et on a

$$\|x_{k+1} - \bar{x}\| \leq C \|x_k - \bar{x}\|^2.$$

La vitesse de convergence de l'algorithme de Newton permet de dire que le nombre de chiffres décimaux calculé exactement double à chaque itération, contre ce qui se passerait si on avait $\|x_k - \bar{x}\| \leq C\lambda^k$ où il faudrait un nombre fixe d'itérations pour l'augmenter de 1. Par contre, la convergence demande à partir d'un point qui soit déjà suffisamment proche de \bar{x} . Aussi, il est nécessaire à chaque itération de l'algorithme d'inverser la différentielle $DF(x_k)$ ce qui, en pratique, peut assez coûteux computationnellement.

Cela peut s'appliquer pour résoudre $\nabla f = 0$ et on a donc

Théorème 3.15. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction C^3 . Si \bar{x} est un point de minimum de f et $D^2(\bar{x})$ est définie positive, alors il existe un rayon $R > 0$ tel que pour tout point $x_0 \in B(\bar{x}, R)$ la suite donnée par $x_{k+1} = x_k - (DF(x_k))^{-1} F(x_k)$ converge vers \bar{x} et satisfait $\|x_{k+1} - \bar{x}\| \leq C \|x_k - \bar{x}\|^2$.*

On remarque que appliquer l'algorithme de Newton à la minimisation de la fonction f correspond à minimiser à chaque itération la fonction

$$\tilde{f}_{x_k}(x) := f(x_k) + \nabla f(x_k) \cdot (x - x_k) + \frac{1}{2} D^2 f(x_k)(x - x_k, x - x_k)$$

qui est le DL_2 de f en x_k . Au contraire, appliquer l'algorithme du gradient à pas fixe correspond à minimiser à chaque itération la fonction

$$\hat{f}_{x_k}(x) := f(x_k) + \nabla f(x_k) \cdot (x - x_k) + \frac{1}{2\tau} \|x - x_k\|^2,$$

qui est une approximation plus grossière de f (mais demande à inverser la matrice identité au lieu de la matrice $D^2f(x_k)$).

L'algorithme de Newton pourrait aussi être utilisé pour résoudre numériquement le système des extrema liés

$$\begin{cases} \nabla f(x) = \sum_{i=1}^k \lambda_i \nabla g_i(x), \\ g_i(x) = 0 \end{cases},$$

en prenant $F : \mathbb{R}^{n+k} \rightarrow \mathbb{R}^{n+k}$ donnée par $F(x, \lambda) = (\nabla f(x) - \sum_i \lambda_i \nabla g_i(x), g(x))$.